FLAME: Free-form Language-based Motion Synthesis & Editing

Abstract

Text-based motion generation models are drawing a surge of interest for their potential for automating the motion-making process in the game, animation, or robot industries. In this paper, we propose a diffusion-based motion synthesis and editing model named FLAME. Inspired by the recent successes in diffusion models, we integrate diffusion-based generative models into the motion domain. FLAME can generate high-fidelity motions well aligned with the given text. Also, it can edit the parts of the motion, both frame-wise and joint-wise, without any fine-tuning. FLAME involves a new transformer-based architecture we devise to better handle motion data, which is found to be crucial to manage variable-length motions and well attend to free-form text. In experiments, we show that FLAME achieves state-of-the-art generation performances on three text-motion datasets: HumanML3D, BABEL, and KIT. We also demonstrate that FLAME's editing capability can be extended to other tasks such as motion prediction or motion in-betweening, which have been previously covered by dedicated models.